Using Jetstream to Enable Large-Scale Text Analysis of Tweets*

Harrison Wittenbrook
Kennesaw State University
Kennesaw, Georgia
hwittenb@students.kennesaw.edu

Lorissa Humble
New Mexico State University
Las Cruces, New Mexico
lorissah@nmsu.edu

David Kloster
Indiana University
Bloomington, Indiana
klosteda@iu.edu

Tassie Gniady
Indiana University
Bloomington, Indiana
ctgniady@iu.edu

ABSTRACT

Performing large-scale text analysis is becoming increasingly common across a wide variety of research domains — from computer science to the humanities to psychology. The aim of large-scale text analysis is to reveal patterns which would not be readily discernible by a human alone. One of the tools of text analysis which is becoming increasingly popular is sentiment analysis. Generally, in sentiment analysis, scores (either positive or negative) are assigned to a piece of text in an automatic fashion without the "human in the loop". There are three general sentiment detection methods which have been studied: dictionary-based methods, supervised learning methods, and unsupervised/deep learning methods. In this work, we focus on comparing two dictionary-based methods (VADER, HL) using Twitter data. To do this, we use >2,000,000 Tweets from Indiana University's OSoMe Twitter collection (10% random sample of public tweets going back to August 1, 2016). We store data and process the tweets with each sentiment dictionary on the Jetstream Cloud Computing system. We then compare word counts and similarity measures to assess the dictionary robustness.

CCS CONCEPTS

• **Computer systems organization → cloud computing**; • **Information retrieval → sentiment analysis**; • **Human-centered computing** → social media;

KEYWORDS

Virtual machines, cloud, Twitter, sentiment analysis, data storage

ACM Reference Format:
Harrison Wittenbrook, Lorissa Humble, David Kloster, and Tassie Gniady. 2018. Using Jetstream to Enable Large-Scale Text Analysis of Tweets. In *Proceedings of PEARC Conference (PEARC'18).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

*Produces the permission block, and copyright information

PEARC'18, July 2018, Pittsburgh, PA USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Over the course of President Trump's term in office, relations with Kim Jong-un's North Korean regime have shifted dramatically. Within 18 months, tension between the United States and North Korea deescalated from the possibility of nuclear war to a face-to-face meeting between the two leaders. Have the shifting relations between the two countries influenced the general public's sentiment regarding North Korea?

Through the use of Indiana University's OSoMe API, the IUNI observatory on social media, which collects a random sample of 10% of all tweets posted online, we searched for tweets using the phrase "North Korea" from January 20th, 2017 to June 19th, 2018. Furthermore, we were able to use the VADER sentiment analysis algorithm, which specializes in analyzing social media posts by running the text through its sentiment lexicon, and the polarity dictionary created by Minqing Hu and Bing Liu (HL), to determine the attitude online during that timeframe. As analysis of this magnitude of tweets is a very memory intensive process, we utilized the Jetstream cloud resource [6][7] to perform analysis in a much more efficient manner. We downloaded, stored, and analyzed over 2.6 million tweets using Jetstream's computing capabilities as this project could not be accomplished efficiently using standard hardware.

The data was run with both dictionaries and displayed in line and bar graphs to assess the similarity in patterns of sentiment scores between VADER and HL. In particular, we expect to see a positive turn in opinions regarding North Korea as President Trump began to comment more positively on North Korea and Kim Jong-un.

2 MATERIALS AND METHODS

2.1 Jetstream

Jetstream is a cloud computing resource that enables researchers from a variety of domains to utilize supercomputing power in a user-friendly and remotely accessible format [6]. Jetstream provides access for Extreme Science and Engineering Discovery Environment (XSEDE) researchers to conduct scientific data analysis. Jetstream virtual machines are customizable, allowing researchers to employ a variety of software configurations to fit the purposes of their projects [8]. In this project, we used Jetstream to obtain, store, and analyze Twitter data that standard hardware could not manage due to the large quantity of tweets (2.6 million) and the depth of the analysis performed. We used a Jetstream virtual machine with 120 GB of memory and a CPU with 44 cores.

Table 1: North Korea Data Ranges

Name	Time Period	Number of Tweets	Key Events
Rising Tensions	01/20/2017 — 08/31/2017	807,017	2/11/2017 - First ICBM test
			06/02/2017 - United Nations enact sanctions
			08/28/2017 - North Korea launches missile over Japan
Hostilities	09/01/2017 — 01/08/2018	939,159	09/19/2017 - Trump calls Kim Jong-un "Rocket Man"
			11/19/2017 - North Korean state media announce that they have tested an ICBM that can reach the US
			01/01/2018 - Trump announces that he has a "bigger and more powerful nuclear button, and [his] works"
Conciliatory Efforts	01/09/2018 — 06/19/2018	941,940	02/09/2018 - North and South Korean athletes march together at the Winter Olympics
			04/27/2018 - Kim Jong-un has a meeting with President Moon of South Korea
			06/12/2018 - President Trump and Kim Jong-un have a summit in Singapore

2.2 OSoMe

The Twitter data we used was extracted from the OSoMe data set, which as partially described above, is a repository of 10% of tweets and corresponding metadata from August 1st, 2016 to present [2] Public tweets from the social networking site Twitter are collected via what Twitter assures is a random sampling method–Twitter's Streaming API. The Streaming API establishes a connection between the user's app and delivers data to the user. This resource allowed us to make requests for tweets containing the phrase "North Korea" which created a dataset of 2,688,116 tweets during the period of January 20, 2017 to June 19, 2018. We reasoned that this date range would reflect the dynamics between North Korea and the United States throughout Trump's presidency because it would show how public perception of North Korea is altered as diplomatic relations grow more tense, or contrarily, become increasingly amicable.

2.3 Partitioning Data

Once access to the requested data was granted, we used a Jetstream virtual machine to download and decompress the files containing the tweets. This allowed us to scan samples of the tweets to ensure as much as possible that the majority would be on-topic.

We then separated our Twitter data into three separate date-ranges, which frame the shifting relations between the United States and North Korea, as shown in Table 1. Our first segment began when President Trump was inaugurated on January 20, 2017 and involved a period when the Korean regime expanded their nuclear testing program; during this interval, tensions were on the rise between the two states. The second time-frame defined encompassed heated engagements between the two leaders. This included President Trump famously calling the Korean dictator "little rocket man" and hinting at a possible military conflict. The final segment involved a decline in tensions between the two states as the leaders began speaking more respectfully of one another, and this period concluded on June 19, 2018 following the Singapore summit between Trump and Un on June 12, 2018.

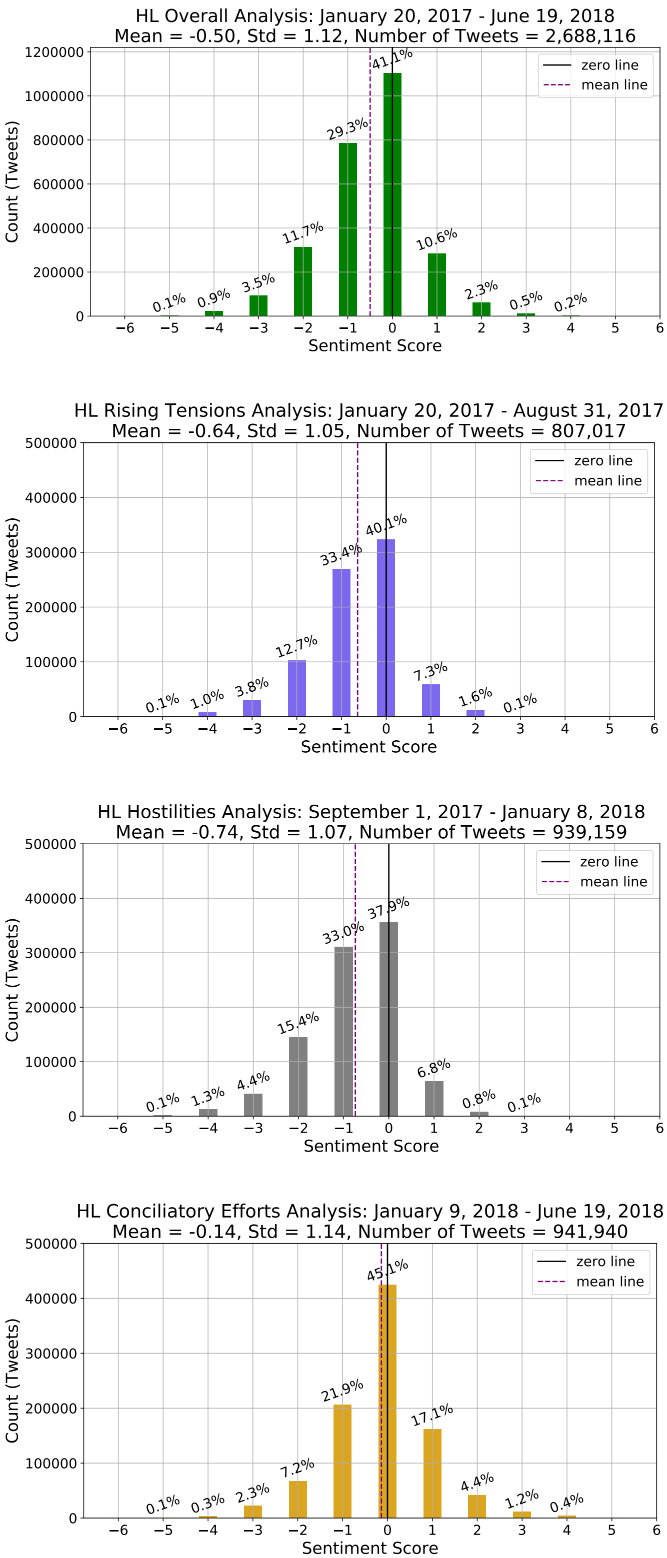
2.4 Sentiment Dictionaries

The Valence Aware Dictionary for sEntiment Reasoning (VADER) [3], a lexicon designed particularly for social media, has the ability to detect the grammatical and syntactical usages of words that express word intensity, word context, word modifiers, punctuation, capitalization, emojis, contrastive conjunctions, and negation. Each word in the rank dictionary is rated from -4 to 4, with zero being neutral. For instance, "magnificently" is given a score of +3.4 and "holocaust" receives a score of -4.

Comparison statements, or sentences that measure one object against another, are also taken into account by VADER. For example, if an individual were to tweet, "the United States is so much better than North Korea," it would be read as a positive statement despite that it detracts from North Korea in order to express a positive opinion towards the United States. Additionally, negation, but-clauses (stating an opinion but using 'but' to indicate a shift or reversal in feeling towards something), and decreasing and increasing words are also detected.

To provide a point of comparison for VADER, the HL polarity dictionary was also employed [9] [4] [5]. In a similar manner to VADER, HL uses a dictionary approach that classifies 'opinion' words into positive, negative, or neutral categories. Words that express something desirable or good about some object, event, or individual possess a positive polarity such as "awesome" or "beautiful," whereas words with negative polarity convey a state that is unwanted such as "gross" or "unfortunate." These words in a given text are matched to the sentiment lexicon so that positive words are given a +1 and negative words are given a -1. Additionally, words that do not appear in either positive or negative dictionary are considered neutral and given no score. The scores are summed to

Figure 1: Intensity of Sentiment Using HL



then reflect the overall rating for the tweet; if both positive and negative words are in a sentence, the same procedure applies.

Because VADER is attuned to social media contexts and registers more complex aspects of language and communication, we hypothesize that VADER will more accurately reflect the sentiment in the tweets. HL records scores on a binary; either the words are positive 1 or negative 1 with no consideration of sentiment intensity ("wonderful" is more positive than "ok" and so on), and the context surrounding the word does not change the actual sentiment being expressed in the tweet. Additionally, HL does not detect n-grams, or words negated by "not." This likely gives its scores a positive skew, even though an individual may be trying to express a negative sentiment, such as "not great." Therefore, VADER is a more nuanced tool that we think will be more accurate given these reasons.

2.5 Modifying Sentiment Dictionaries

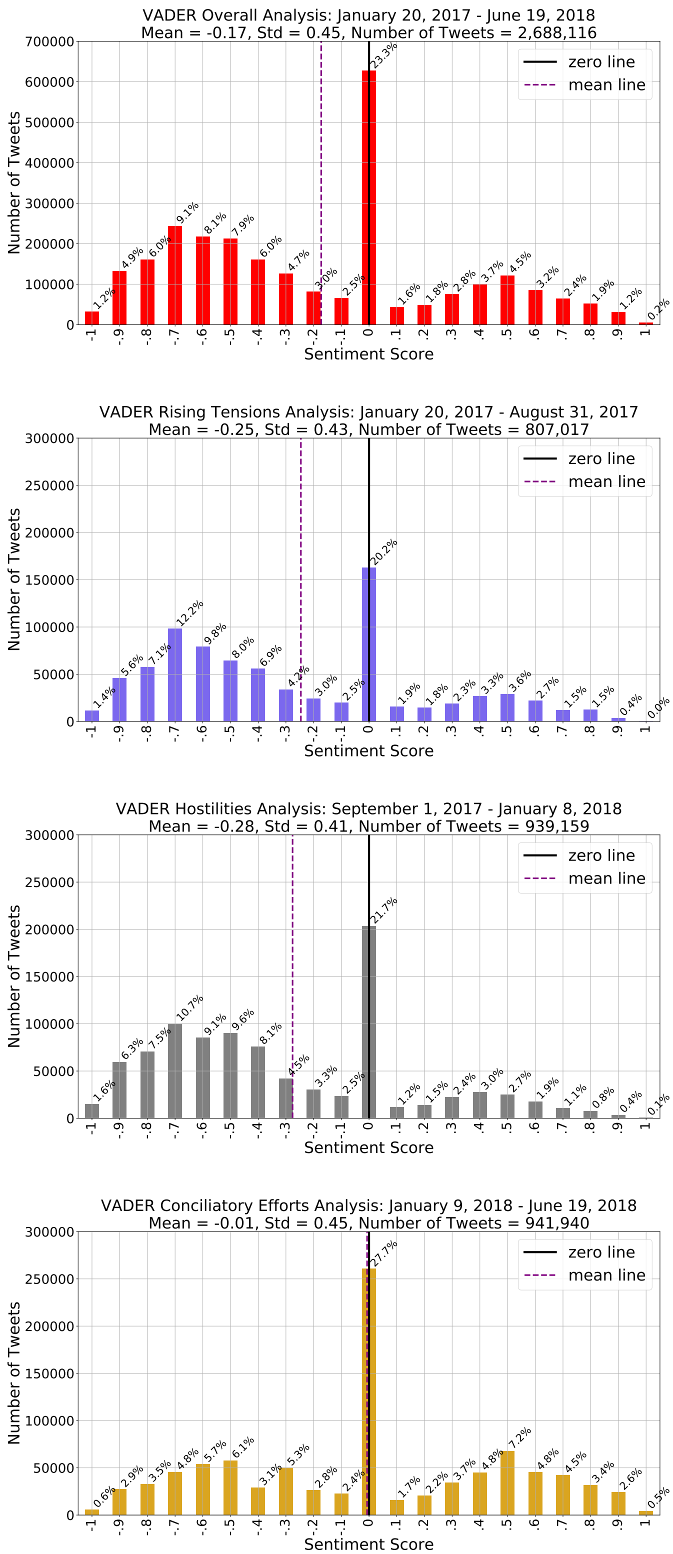
After manually scanning the tweets, we began preliminary analysis. The two dictionaries were used in conjunction with scripts that gave us the one hundred most commonly used positive and negative words in the tweets from each dictionary, with which we were able to modify the sentiment dictionaries to suit this project and to ensure that both the dictionaries were as consistent with each other as possible. We eliminated potentially problematic words such as "like" and "breaking" as they would inaccurately skew the results of the sentiment analysis positively or negatively, respectively; in addition, we added words that previously did not appear in either the positive or negative lists such as "missiles" and "weapons." Occasionally, words that were in the negative or positive category were better suited for the opposite list and were switched. In order to make an adjustment of any kind to the dictionary, the four researchers must have agreed on the change.

Table 2: Words Added to Sentiment Dictionaries

Word	Action Taken
Missile	Added to negative VADER and HL
Nuke	Added to negative VADER and HL
Denuclearization	Added to positive VADER and HL
Sanctions	Added to negative VADER and HL
Shooting	Added to negative VADER and HL
Bombs	Added to negative VADER and HL
Banning	Added to negative VADER and HL
Standoff	Added to negative VADER and HL
Provocation	Added to negative VADER and HL
Decapitation	Added to negative VADER and HL

Adding words to the VADER dictionary required a different process. Each word had to be rated by 10 individuals on a scale of -4 to 4. The raters were asked to score each word as they believed a majority of people would score them in the context of "North Korea." This was done to help reduce personal bias and increase accuracy. The scores given by each rater for a particular word were then averaged to determine its valence score. We added 36 words to the VADER dictionary via this process such as "denuclearize" and

Figure 2: Intensity of Sentiment Using VADER



"ravage." However, adding and removing words from the HL dictionary only required agreement between the researchers. Examples of words added to the valance dictionaries are shown in Table 2.

2.6 Analyzing Data

In order to obtain an overall result for the entire period of January 20th, 2017 to June 19th, 2018, the sentiment scores from the date ranges—Rising Tensions, Hostilities, and Conciliatory Efforts— were combined and run together with the aim of ascertaining the general sentiment towards North Korea throughout Trump's presidency.

We then took the individual ratings received from VADER and HL and graphed the data based on the intensity of the scores. For the HL dictionary, raw scores were taken for each tweet and plotted by date range. Additionally, the mean and standard deviation were calculated for each set of data to help depict the overall trend for that time period. Figure 1 demonstrates the result of this process.

While compiling the data using the HL dictionary gave us whole numbers suitable for this form of graphing, the VADER algorithm returns a floating point number between -1 and 1. Because of this discrepancy, the ratings from VADER were binned based on ranges of scores. For example, a score ranging from $0 < x \leq 0.1$ in the VADER algorithm would be represented as a 0.1. The mean and standard deviation were also calculated for the VADER scores, and Figure 2 represents the results of this analysis for each time period.

After analyzing the results from VADER and HL separately, we began the process of comparing the data from the two dictionaries. We accomplished this by taking a weekly mean of the sentiment scores received from each of the dictionaries and standardized the data using z-scores. We then plotted the results as shown in Figure 3 to demonstrate how each dictionary differs in results over time.

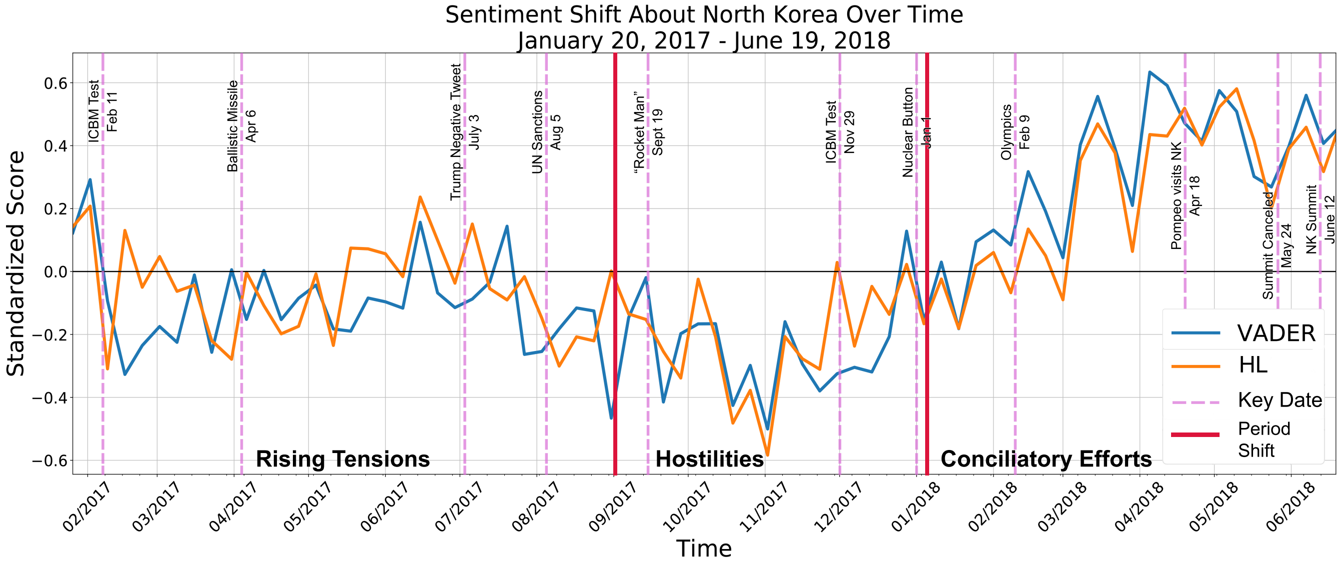
Additionally, to have data independent from President Trump to compare results to, we calculated the mean and standard deviation of scores between the two dictionaries from a control sample taken between August 1, 2016 and December 31, 2016. Our objective in doing this is to establish a baseline of sentiment scores towards North Korea before President Trump took office to determine how much of an effect on public sentiment Donald Trump has had.

3 COMPILING RESULTS

In general, our expectations for the data were confirmed for both the HL polarity dictionary and VADER. The results from the control group can be seen in Table 3 and illustrate the increase in negativity following Trump's inauguration until the most recent date range, Conciliatory Efforts. For the VADER control group, the mean is half as negative as Rising Tensions and Hostilities. It is particularly less negative than HL, with the mean for the control group being half as negative as the two earlier date ranges. As can be seen in the VADER and HL graph in Figure 3, there has been an uptick in positive sentiment regarding North Korea in 2018.

Early in Trump's presidency (the date range of January to August of 2017 termed Rising Tensions) attitudes towards North Korea were largely negative and stayed within the range of 0.3 to -0.35. Continuing to examine Figure 3, North Korea had begun testing missiles and conducted four missile tests in the month of February, which lowered sentiments substantially as registered by VADER.

Figure 3: Z-scored sentiment shift over time using VADER and HL



Notably, during March VADER and HL diverged, with the HL algorithm showing an upturn in positivity. Although still negative, April for both VADER and HL saw a slight shift towards neutrality followed by a further increase in positivity in May despite events that were expected to cause a decrease such as the ballistic missile launch and warnings from the U.S. security council for North Korea to cease missile tests. However, in May, President Trump admitted that Kim Jong-un was a "smart cookie" and expressed his desire to meet with the North Korean leader, which coincided with a spike in positivity for both HL and VADER. The missile launch conducted by Kim Jong-un that same month and the imposition of sanctions on North Korea were expected to lower positivity, yet it remained more positive than expected, especially for HL, which jumped above the mean line in mid-May. June reached a peak high for this date range despite the fact that North Korea continued to test missiles, including the instances where missiles were fired into the sea near the Korean peninsula. Late July began to grow increasingly negative, possibly due to Trump reacting with the tweet, "North Korea just launched another missile. Does this guy have nothing better to do with his life?" Not surprisingly, following the missile launch that flew over Japan in August, VADER and HL have a drastic shift in opinion towards the negative end of the sentiment scale. In August, additional sanctions were imposed on North Korea and regardless of this the country continued to test missiles, later threatening to target Guam. Trump's remark that he would meet North Korea with "fire and fury like the world has never seen" likely exacerbated the negative attitudes toward the country.

September 2017 through early January 2018 exhibited a large downturn in sentiment as the interactions between Trump and Kim became more hostile and inflammatory. More extreme threats were exchanged between the two, such as Trump expressing his willingness to "totally destroy" North Korea and Kim contending that he will "tame the mentally deranged dotard with fire." In both HL and VADER, November reached a new low in public sentiment as the contentious interactions continued. As expected, this date range (labelled Hostilities) was the most negative in terms of sentiments.

In January 2018, positivity improved but was rather inconsistent. Trump and Kim's interactions were a mix of virulence and expressions of the desire to meet one another and come to an agreement. Particularly as the Winter Olympics approached in February and South and North Korean athletes marched together in the opening ceremony, there was a resurgence in positivity. In March, which saw a nearly continuous increase in favorable attitudes, Kim suggested that he would be willing to discuss the North Korean nuclear arsenal. In the middle of March, however, the additional missile tests conducted by North Korea likely prompted a downswing in sentiments. Following this in April, Mike Pompeo and Kim Jong-un met and, according to Trump, "cultivate[d] a good relationship," and in addition to this, North Korea suspended missile tests and released some political prisoners. As the events in April were mostly positive, so was public sentiment. However, even as May brought promises of a summit and optimism from Trump about the potential for peace, the middle of May saw a downturn in sentiment following the cancellation of the meeting. Finally, after the summit was announced to take place after all, Trump and Kim met on June 12th, prompting an uptick in positivity. VADER and HL, particularly for this last date range, began to more closely resemble each other's sentiment scores and follow the same general trends.

Overall, the general intensity of sentiment used shifted over the three date ranges observed. As shown in Figure 2, sentiment became slightly more vehement between the Rising Tensions and Hostilities periods when measured using the VADER lexicon. However, in the period of Conciliatory Efforts, sentiment intensity shifted drastically by becoming far more neutral overall with the positive and negative sentiment being largely balanced. A similar development is expressed in Figure 1, which shows the sentiment intensity for the HL dictionary. While the shifts in severity of sentiment are less drastic for the HL dictionary, similar patterns to the VADER dictionary are followed as the sentiment increases in negativity then balances out.

Table 3: Results of Statistical Tests

	Time Frame	Mean	Z-Score Mean	SD	P Value[a]
VADER	Before Inauguration	-0.12	0.12	0.42	-
	Rising Tensions	-0.25	-0.16	0.43	< 0.001
	Hostilities	-0.28	-0.23	0.40	
	Conciliatory Efforts	-0.01	0.37	0.45	
HL	Before Inauguration	-0.38	0.10	0.95	-
	Rising Tensions	-0.64	-0.13	1.05	< 0.001
	Hostilities	-0.74	-0.22	1.07	
	Conciliatory Efforts	-0.14	0.32	1.14	

[a]For VADER we used the Kruskal-Wallis test, and for HL we used the ANOVA with Contrasts test

As stated above, we are interested in whether there has been a significant shift in public perception of North Korea throughout Trump's presidency. The relations between the United States and North Korea have progressed through a series of stages that we framed in the date ranges from Table 1. To test this assumption statistically we conducted an analysis of variance (ANOVA) for both the HL dictionary results and the VADER results with the goal of determining whether there are statistically significant differences between the three date ranges, Rising Tensions, Hostilities, and Conciliatory efforts. The results are shown in Table 3 along with results from the control group for comparison. Since ANOVA is an omnibus test and does not specify which group is different, we also ran an ANOVA with contrasts in order to compare specific pairs of date ranges to test for significance. The Kruskal-Wallis nonparametric test was used for the VADER distribution of scores as it was not a normal distribution. Upon initially inspecting the bar graphs of sentiment scores as well as the means and standard deviations for the sentiment scores of each date range, it could be seen that the date range of Conciliatory Efforts consisted of more positive and neutral tweets than either Hostilities and Rising Tensions; Hostilities had more negative tweets than Rising Tensions.

For the VADER results, the Kruskal-Wallis test for differences between the three date ranges was significant at 0.001 with a p-value of $2.2e - 16$. Rising Tensions had a mean sentiment score of -0.25, Hostilities had a mean of -0.28, and Conciliatory Efforts was -0.01. This result alongside the means suggests that there are significant differences between date ranges in terms of sentiment scores, with the last date range being more positive than the other two. Since ANOVA with contrasts is a parametric test it could not be paired with the Kruskal-Wallis test. However, for the HL results we were able to conduct an ANOVA with contrasts which yielded a significant result at 0.001 with a p-value of $2e - 16$. The mean sentiment scores for HL were -0.64, -0.74, and -0.14 respectively. The groups we compared for the planned contrasts were the sentiment scores for Rising Tensions to Hostilities, Hostilities to Conciliatory Efforts, and Conciliatory Efforts to Rising Tensions; all comparisons were significant with a p-value of $2e - 16$, meaning that they are significantly different from each other. Examination of the means suggests that the direction of the difference between the first pair is negative, whereas the latter pairs has a positive difference. Therefore, the most recent date range, Conciliatory Efforts, is significantly more positive than the former date ranges. Likewise, Hostilities is significantly more negative than Rising Tensions.

4 DISCUSSION

This research has attempted to use the VADER sentiment analysis tool and the HL polarity dictionary to gauge public sentiment towards North Korea throughout Trump's presidency thus far. Although it is by no means a perfect method, it yielded interesting results that highlight the usefulness of these tools and text analysis in general to reveal patterns in public perception towards significant events. Sentiment analysis of tweets in particular has a unique ability to do this as it is through social media that people have a public platform to express their ideas and opinions.

The results discussed above demonstrated that sentiment towards North Korea is variable and depends on events in the relations between the representatives of the United States and North Korea. This is not to diminish the effect that North Korea's actions, including the missile tests and human rights violations, had on people's opinions of the country. The sentiment scores generated around the time of those incidents showed that this indeed had an influence on how people regarded North Korea. However, we would like to make the point that the manner in which the President interacted with Kim Jong-un fundamentally impacted how people perceived the situation. The date ranges we chose to frame the types of interactions Trump had with Kim Jong-un were intended to capture this effect.

Specifically, and similarly to what we posited earlier, President Trump's opinions about North Korea and Kim Jong-un in particular have ranged from extremely deprecatory insults and even threats to destroy the country through nuclear warfare to complimenting the North Korean leader's intelligence and efforts to establish peace. Exactly how much of an effect these relations and publicized interactions have had on people's opinion towards North Korea is difficult to precisely ascertain, but our analyses give us a good estimate. The results from the ANOVA and Kruskal-Wallis as well as the graphs printed above partially illustrate the change in sentiments over time in conjunction with key events in the relations between the United States and North Korea. This illuminates the relationship between the actions and words of a country's leaders and the attitudes and beliefs of its citizens; when an individual is in a position of authority such as this, there is the potential for them to have considerable sway over the populace. Conservatively, however, it would be justified to assert that Trump's behavior and his use of Twitter as a medium to communicate his opinions has had at least a partial influence over American's attitudes towards North Korea.

This is also illustrated in the work of George Lakoff who has conducted research on Donald Trump's tweets and has described how the President has employed social media to influence public perception [1]. He has categorized Trump's tweets into four classes: preemptive framing, diversion, deflection, and trial balloon. Firstly, Trump's preemptive framing tweets are used to control the narrative on an issue before mainstream media has the opportunity to put forth an official account. Secondly, diversion is used to direct attention away from more important issues by introducing through his Twitter feed more trivial matters. Thirdly, Trump uses deflection to change the direction of a situation that is not working in his favor by insulting the other party. Lastly, trial balloon tweets are posted in order to determine the public's attitudes on a certain issue. His tweets have become his favored mode of communication to reach people and influence the news cycle, as his tweets receive attention from traditional media in addition to the people on Twitter who see what he posts.

In using both VADER and the HL polarity dictionary, we were able to compare two well-known sentiment analysis tools and obtained a better understanding of the data as a result. The conclusion that can be reached from this is that when conducting sentiment analysis, it is most appropriate to use more than one tool. This allows a researcher to check for the accuracy of their results and examine inconsistencies in sentiment dictionaries and sentiment scores. We were often able to fill in gaps in the positive and negative words in either dictionary by looking at the other's analysis of the text, adding or removing certain words from either dictionary as they appeared in the positive and negative word frequencies that were run before the official analyses.

VADER, as it was designed for social media and detects many facets of language and communication, was a more nuanced tool than the polarity dictionary alone. It thus produced a more accurate reflection of the attitudes and evaluations present in the tweets and compensated for the polarity dictionary's deficits. However, this is not to say that important information was not rendered by the polarity dictionary. In fact, we believe that it was complimentary to VADER in that it registered words that VADER may have missed.

Our ability to examine this data is owed to Jetstream, which provided the computing resources needed to manage and analyze large quantities of rich data in order to answer complex questions about public perception. Using Jetstream's cloud computing system, the 2.6 million tweets we collected from the OSoMe data set could be stored, displayed, and processed through the virtual machines that this resource makes accessible to researchers. In addition, due to Jetstream's customizability, we could choose our virtual machine image based on the needs of our project and add the necessary software to the image, such as Jupyter Notebooks and RStudio for writing and compiling code.

4.1 Limitations

Throughout the process of analyzing the tweets surrounding North Korea, our team encountered limitations which interfered with our research. One of the most drastic of these limitations was the lack of certainty that data collected from the Twitter API was a random sample of all tweets posted. If tweets were selected by Twitter by non-random means, the results from our sentiment analysis could be skewed. Additionally, another limitation from the Twitter data collection process was OSoMe's ability to only collect 10% of tweets posted online. By only having a partial sample of the overall content, it can be difficult to say with certainty that our results reflect the overall sentiment surrounding North Korea. Furthermore, the three date ranges selected — Rising Tensions, Hostilities, and Conciliatory Efforts — were chosen based on potentially impactful events in the North Korean timeline by the researchers. This process was mostly subjective and lead to slightly unequal categories.

4.2 Possible Directions for Future Research

While our research into the shift in attitudes toward North Korea through Twitter sentiment analysis did present interesting findings, there are other approaches to sentiment analysis which could yield fruitful results in later projects. One approach to further research would be a comparison of additional dictionaries. Using sentiment dictionaries designed for different purposes to analyze text would be a pertinent step to take following this research.

Another possible avenue of future research could involve employing different methods of sentiment analysis. One commonly used method is using machine learning to score tweets rather than usage of a lexicon. By pursuing this method of rating tweets, results could be compared to those of a lexicon approach to determine any significant changes between the two methods.

Because we are working with such large quantities of data, certain issues with analysis arise such as the tendency for tests to reach significance for even minor or potentially meaningless differences. Whether the differences detected by the statistical test are meaningful differences is more difficult to determine. Future research should use Bayesian statistics which compensates for the problems that come with data analytics with big data.

5 CONCLUSION

The purpose of this project has been to uncover patterns in public perception of events in North Korea as Trump and Kim Jong-un's diplomatic relations with one another have unfolded, which conveys the extent to which a country's leaders can influence the attitudes and opinions of the populace. Throughout Trump's presidency, his interactions with Kim Jong-un have ranged from peace talks to threats of nuclear annihilation, and he has communicated his attitudes towards the North Korean state quite publicly via Twitter or press conferences. As these events have occurred, public sentiment changed accordingly, often to match the attitudes expressed by the President. This calls attention to how a leader's words and actions can have a large impact on the way people perceive issues—thus highlighting the importance of ensuring that the information the President presents to the public is accurate. The fact that there are significant differences between the 3 time periods shows that the kinds of interactions that Trump has had with Kim Jong-un have altered the way that the American public evaluates the situation at hand as the variation in Trump's attitudes towards North Korea coincided with changes in public sentiment. Large-scale text and sentiment analysis, enabled by Jetstream, has proven itself to be a valuable method to examine and measure social patterns and change, and would be useful in many other scenarios to understand the complexity of social life amongst other applications.

REFERENCES

[1] Andrew Buncombe. 2018. Donald Trump one year on: How the Twitter President changed social media and the country's top office. *Independent* (Jan 2018). https://independent.co.uk/news/world/americas/us-politics/the-twitter-president-how-potus-changed-social-media-and-the-presidency-a8164161.html

[2] Clayton A Davis, Giovanni Luca Ciampaglia, Luca Maria Aiello, Keychul Chung, Michael D Conover, Emilio Ferrara, Alessandro Flammini, Geoffrey C Fox, Xiaoming Gao, Bruno Gonçalves, Przemyslaw A Grabowicz, Kibeom Hong, Pik-Mai Hui, Scott McCaulay, Karissa McKelvey, Mark R Meiss, Snehal Patil, Chathuri Peli Kankanamalage, Valentin Pentchev, Judy Qiu, Jacob Ratkiewicz, Alex Rudnick, Benjamin Serrette, Prashant Shiralkar, Onur Varol, Lilian Weng, Tak-Lon Wu, Andrew J Younge, and Filippo Menczer. 2016. OSoMe: The IUNI observatory on social media. *PeerJ Computer Science* 2 (2016), e87. https://doi.org/10.7717/peerj-cs.87

[3] CJ Hutto Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf.*

[4] Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 168–177.

[5] Bing Liu. 2010. Sentiment Analysis and Subjectivity. *Handbook of natural language processing* 2 (2010), 627–666.

[6] Craig A. Stewart, Timothy M. Cockerill, Ian Foster, David Hancock, Nirav Merchant, Edwin Skidmore, Daniel Stanzione, James Taylor, Steven Tuecke, George Turner, Matthew Vaughn, and Niall I. Gaffney. 2015. Jetstream: A Self-provisioned, Scalable Science and Engineering Cloud Environment. In *Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure (XSEDE '15)*. ACM, New York, NY, USA, Article 29, 8 pages. https://doi.org/10.1145/2792745.2792774

[7] Craig A. Stewart, David Y. Hancock, Matthew Vaughn, Jeremy Fischer, Tim Cockerill, Lee Liming, Nirav Merchant, Therese Miller, John Michael Lowe, Daniel C. Stanzione, James Taylor, and Edwin Skidmore. 2016. Jetstream: Performance, Early Experiences, and Early Results. In *Proceedings of the XSEDE16 Conference on Diversity, Big Data, and Science at Scale (XSEDE16)*. ACM, New York, NY, USA, Article 22, 8 pages. https://doi.org/10.1145/2949550.2949639

[8] J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott, and N. Wilkins-Diehr. 2014. XSEDE: Accelerating Scientific Discovery. *Computing in Science Engineering* 16, 5 (Sept.-Oct. 2014), 62–74. https://doi.org/10.1109/MCSE.2014.80

[9] Ley Zhang, Riddhiman Ghosh, Mohamed Dekhil, Meichun Hsu, and Bing Liu. 2011. Combining lexicon-based and learning-based methods for Twitter sentiment analysis. *HP Laboratories, Technical Report HPL-2011* 89 (2011).